

A STOCHASTIC RELIABILITY MODEL OF A SERVER UNDER A RANDOM WORKLOAD

RACHEL TRAYLOR,* *EMC, University of Texas at Arlington*

Abstract

Traffic to any server is rarely constant over time. In addition, the workload brought by each service request is typically unknown in advance, and each request may bring a different workload to the server. Cha and Lee (2011) proposed a reliability model where each request brings an identical and constant workload. In this paper, we generalize the model to allow for requests to bring an unknown random stress to the server. Jobs arrive to the server via a nonhomogeneous Poisson process. Service times are random and i.i.d. Each job adds a random stress $\mathcal{H}_j \sim \mathcal{H}$ to the breakdown rate of the server until the job is completed. The survival function of such a server and the efficiency of the server are derived.

Keywords: Stochastic Processes; Nonhomogeneous Poisson process; Traffic Modeling; Web server

2010 Mathematics Subject Classification: Primary 60G20

Secondary

1. Introduction

A stochastic model for server reliability under nonconstant traffic has many applications including retail, logistics, manufacturing, and IT. In [1], the authors considered a Web server wherein each request arrives via a nonhomogeneous Poisson process $\{N(t) : t \geq 0\}$ with intensity $\lambda(t)$ and adds a constant stress η as long as the job is in the system. Assuming the baseline breakdown rate is $r_0(t)$, the breakdown rate process is defined as

$$\mathcal{B}(t) = r_0(t) + \eta \sum_{j=1}^{N(t)} \mathbb{I}(T_j \leq t \leq T_j + W_j)$$

where $N(t)$ is the nonhomogeneous Poisson process describing the arrivals of requests, $\{T_j\}_{j=1}^{N(t)}$ are the arrival times of requests, $\{W_j\}_{j=1}^{N(t)}$ are the service times, and η is the constant stress breakdown rate (CSBR) increase brought by each job. It is assumed that the $\{T_j\}_{j=1}^{N(t)}$ are independent and the $\{W_j\}_{j=1}^{N(t)}$

* Postal address: 2421 Mission College Blvd. Santa Clara, CA 95050

are i.i.d. with pdf $g_W(w)$ and distribution $G_W(w)$. Cha and Lee (2011) derive the survival function for such a server.

The assumption that each job brings the same stress to a server, especially a Web server, does not reflect applications accurately. In many situations, each job will bring a different, unknown stress to the server, in addition to a random arrival and service time. Therefore, the following generalization and expansion of the model is proposed. It should be noted that the model is general enough to apply to any situation with a server or system that experiences nonconstant traffic with unknown and random workloads, or stresses, brought by each service request.

2. Random Stress Breakdown Rate

Retaining the previous assumptions from [1] except for the supposition that every request adds a constant stress to the server, the following generalization is proposed. Assume that each job j adds a random stress \mathcal{H}_j to the server for the duration of its time in the system. Suppose the stresses $\{\mathcal{H}_j\}_{j=1}^{N(t)} \stackrel{i.i.d.}{\sim} \mathcal{H}$, where WLOG \mathcal{H} is a discrete random variable with a finite sample space $S = \{\eta_1, \dots, \eta_m\}$ and probability distribution given by $P(\mathcal{H} = \eta_i) = p_i, i = 1, \dots, m, \sum_{i=1}^m p_i = 1$. (Note, the proof is unaffected by the discreteness of \mathcal{H} . One may also take to be a continuous random variable.) Then the new breakdown rate process is given by

$$\mathcal{B}(t) = r_0(t) + \sum_{j=1}^{N(t)} \mathcal{H}_j \mathbb{I}(T_j \leq t \leq W_j) \quad (2.1)$$

Let Y be the random time to failure of the Web server under the workload from client requests. Then the survival function of the server, conditioned on arrival process of client requests ($N(t)$), arrival times ($\{T_j\}_{j=1}^{N(t)}$), service times ($\{W_j\}_{j=1}^{N(t)}$), and stresses ($\{\mathcal{H}_j\}_{j=1}^{N(t)}$) is given by

$$\begin{aligned} P(Y > t \mid N(t), \{T_j\}_{j=1}^{N(t)}, \{W_j\}_{j=1}^{N(t)}, \{\mathcal{H}_j\}_{j=1}^{N(t)}) &= \exp\left(-\int_0^t \mathcal{B}(t) dt\right) \\ &= \bar{F}_0(t) \exp\left(-\sum_{j=1}^{N(t)} \mathcal{H}_j \min(W_j, t - T_j)\right) \end{aligned} \quad (2.2)$$

where $\bar{F}_0(t) = \exp\left(-\int_0^t r_0(x) dx\right)$.

3. Survival Function to RSBR Server

The unconditional survival function $S_Y(t)$ of the Web server under the random stress breakdown rate (RSBR) model is given in the following theorem:

Theorem 3.1. (RSBR Survival Function.) *Suppose that jobs arrive to a server according to a nonhomogeneous Poisson process $\{N(t), t \geq 0\}$ with intensity function $\lambda(t)$ and denote $m(t) \equiv E[N(t)] = \int_0^t \lambda(s) ds$,*

where $m(t)$ has an inverse. Let the arrival times $\{T_j\}_{j=1}^{N(t)}$ be independent. Let the service times $\{W_j\}_{j=1}^{N(t)}$ be i.i.d. with pdf $g_W(w)$ and distribution $G_W(w)$, and let them be mutually independent of all arrival times. Let the stresses $\{\mathcal{H}_j\}_{j=1}^{N(t)} \stackrel{i.i.d.}{\sim} \mathcal{H}$, where the distribution of \mathcal{H} is given above WLOG. Then

$$S_Y(t) = \bar{F}_0(t) \exp \left(-E_{\mathcal{H}} \left[\mathcal{H} \int_0^t e^{-\mathcal{H}w} m(t-w) \bar{G}_W(w) dw \right] \right) \quad (3.1)$$

Proof.

$$\begin{aligned} S_Y(t) &= P(Y > t) \\ &= E \left[P \left(Y > t | N(t), \{T_j\}_{j=1}^{N(t)}, \{W_j\}_{j=1}^{N(t)}, \{\mathcal{H}_j\}_{j=1}^{N(t)} \right) \right] \end{aligned}$$

Using the Law of Total Expectation,

$$E \left[\exp \left(- \sum_{j=1}^{N(t)} \mathcal{H}_j \min(W_j, t - T_j) \right) \right] = E \left[E \left[\exp \left(- \sum_{j=1}^{N(t)} \mathcal{H}_j \min(W_j, t - T_j) \right) \middle| N(t), \{\mathcal{H}_j\}_{j=1}^{N(t)} \right] \right]$$

Let $N(t) = n$ and $\mathcal{H}_j = \eta_{i_j}$ for some $i \in \{1, \dots, m\}$, $j = 1, \dots, n$. Since the stresses brought by each job are mutually independent of the arrival times, $f_{T_1, \dots, T_n | N(t), \mathcal{H}_1, \dots, \mathcal{H}_n}(t_1, \dots, t_n | n, \eta_{i_1}, \dots, \eta_{i_n}) = f_{T_1, \dots, T_n | N(t)}(t_1, \dots, t_n | n)$. Let T'_1, \dots, T'_n be i.i.d. random variables with pdf $f(x) = \frac{\lambda(x)}{m(t)}$. Then by Lemma A.1, the conditional joint distribution of T_1, \dots, T_n is equal to the distribution of $T'_{[1]}, \dots, T'_{[n]}$, where $\{T'_{[i]}\}_{i=1}^n$ are the order statistics of T'_1, \dots, T'_n and is given by

$$f_{T_1, \dots, T_n | N(t)}(t_1, \dots, t_n | n) = n! \prod_{j=1}^n \frac{\lambda(t_j)}{m(t)}, \quad 0 \leq t_1 \leq \dots \leq t_n \leq t$$

Then

$$\begin{aligned} E \left[\exp \left(- \sum_{j=1}^{N(t)} \mathcal{H}_j \min(W_j, t - T_j) \right) \middle| N(t), \mathcal{H}_1, \dots, \mathcal{H}_{N(t)} \right] &= E \left[\exp \left(- \sum_{j=1}^n \eta_{i_j} \min(W_j, t - T_j) \right) \right] \\ &= E \left[\exp \left(- \sum_{j=1}^n \eta_{i_{[j]'}} \min(W_j, t - T_{(j)}) \right) \right] \\ &= E \left[\exp \left(- \sum_{j=1}^n \eta_{i_{j'}} \min(W_j, t - T_{j'}) \right) \right] \\ &= E \left[\prod_{j=1}^n \exp \left(- \eta_{i_{j'}} \min(W_j, t - T_{j'}) \right) \right] \\ &= \prod_{j=1}^n E \left[\exp \left(- \eta_{i_{j'}} \min(W_j, t - T_{j'}) \right) \right] \end{aligned} \quad (3.2)$$

Where the equalities hold because $\{W_j\}_1^n, \{T_{j'}\}_1^n$ are i.i.d., respectively, and mutually independent. Now, fix j' . Then $\eta_{i_{j'}}$ is also fixed, and

$$\begin{aligned} E \left[\exp \left(-\eta_{i_{j'}} \min(W_j, t - T_{j'}) \right) \right] &= E \left[E \left[-\eta_{i_{j'}} \min(W_j, t - T_{j'}) \middle| W_j \right] \right] \\ &= \frac{1}{m(t)} \left(m(t) - \eta_{i_{j'}} \int_0^t \exp(-\eta_{i_{j'}} w) m(t-w) \bar{G}_W(w) dw \right) \end{aligned} \quad (3.3)$$

by A.2. But Equation (3.3) is true $\forall j$, so

$$E \left[\exp \left(- \sum_{j=1}^{N(t)} \mathcal{H}_j \min(W_j, t - T_j) \right) \middle| N(t), \mathcal{H}_1, \dots, \mathcal{H}_{N(t)} \right] = \prod_{j=1}^n \frac{1}{m(t)} \left(m(t) - \eta_{j'}^i \int_0^t e^{-\eta_{j'}^i w} m(t-w) \bar{G}_W(w) dw \right) \quad (3.4)$$

Now it remains to take the expectation of (3.4) over $N(t), \mathcal{H}_1, \dots, \mathcal{H}_{N(t)}$. Let $h_j^{ij}(t) = \left(m(t) - \eta_j^i \int_0^t \exp(-\eta_j^i w) m(t-w) \bar{G}_W(w) dw \right)$. Then

$$E \left[\prod_{j=1}^n \frac{1}{m(t)} \left(m(t) - \eta_{i_j} \int_0^t e^{-\eta_{i_j} w} m(t-w) \bar{G}_W(w) dw \right) \right] = E \left[\prod_{j=1}^n \frac{1}{m(t)} h_j^{i_j}(t) \right]$$

and hence

$$\begin{aligned} E \left[\prod_{j=1}^n \frac{1}{m(t)} h_{i_j}(t) \right] &= \sum_{n=0}^{\infty} \sum_{i_n=1}^m \cdots \sum_{i_2=1}^m \sum_{i_1=1}^m \left(\prod_{j=1}^n \frac{1}{m(t)} h_{i_j}(t) \right) P(\mathcal{H}_1 = \eta_{i_1}) P(\mathcal{H}_2 = \eta_{i_2}) \cdots P(\mathcal{H}_n = \eta_{i_n}) P(N(t) = n) \\ &= \sum_{n=0}^{\infty} \frac{1}{m(t)^n} (E_{\mathcal{H}_1}[h_{i_1}(t)]) \sum_{i_n=1}^m \cdots \sum_{i_2=1}^m \left(\prod_{j=2}^n h_{i_j}(t) \right) P(\mathcal{H}_2 = \eta_{i_2}) \cdots P(\mathcal{H}_n = \eta_{i_n}) P(N(t) = n) \\ &= \sum_{n=0}^{\infty} \frac{1}{m(t)^n} \left(\prod_{j=1}^n E_{\mathcal{H}_j}[h_{i_j}(t)] \right) P(N(t) = n) \\ &= \sum_{n=0}^{\infty} \frac{1}{m(t)^n} \left(\prod_{j=1}^n E_{\mathcal{H}_j}[h_{i_j}(t)] \right) \frac{m(t)^n}{n!} e^{-m(t)} \end{aligned}$$

Since $\mathcal{H}_j \stackrel{i.i.d.}{\sim} \mathcal{H} \forall j$,

$$\begin{aligned} \sum_{n=0}^{\infty} \prod_{j=1}^n E_{\mathcal{H}_j}[h_{i_j}(t)] \frac{e^{-m(t)}}{n!} &= \sum_{n=0}^{\infty} \frac{e^{-m(t)}}{n!} (E_{\mathcal{H}}[h_j(t)])^n \\ &= \exp \left(-E_{\mathcal{H}} \left[\mathcal{H} \int_0^t e^{-\mathcal{H} w} m(t-w) \bar{G}_W(w) dw \right] \right) \end{aligned}$$

The hazard function $r(t)$ is then given by

$$r(t) = -\frac{d \ln(S_Y(t))}{dt} = r_0(t) + E_{\mathcal{H}}[\mathcal{H} \int_0^t e^{-\mathcal{H} w} m(t-w) \bar{G}_W(w) dw] \quad (3.5)$$

4. Efficiency Measure of the Server under RSBR model

If the server crashes during operation, then it is rebooted. We assume that after rebooting, the server is "as good as new", i.e. upon fixing malfunctions, the performance can be regarded as the same as prior to the crash. Cha and Lee (2011) assumed that the arrival process of the client requests after rebooting "restarts". Formally, that

- (1) The arrival process after rebooting, $\{N_{rb}(t), t \geq 0\}$, remains a nonhomogeneous Poisson process with the same intensity function $\lambda(t)$ as before, and
- (2) $\{N_{rb}(t), t \geq 0\}$ is independent of $\{N(t), t \geq 0\}$.

Hence, after a reboot, we may refer to $\{N_{rb}(t), t \geq 0\}$ as $\{N(t), t \geq 0\}$. In addition, assume the time to reboot of the server follows any continuous distribution $H(t)$ with a mean ν , as in [1].

Let $M(t)$ be the total number of jobs completed by the server during $(0, t]$. We wish to measure the performance of the server at any time t , and so retain the definition given in [1].

Definition.[Efficiency of the Server] The *efficiency* of the server, ψ , is defined as

$$\psi := \lim_{t \rightarrow \infty} \frac{E[M(t)]}{t}$$

Standard renewal theory gives $\psi = \frac{E[M]}{E[Y] + \nu}$ where M is the number of jobs completed in a renewal cycle, Y is the length of a renewal cycle, and ν is the mean time to reboot [2]. Under the RSBR model, ψ has a closed form given in the following theorem.

Theorem 4.1. (Efficiency of a Server under RSBR.) *Let the mean time to reboot of the server be ν . Under the same conditions as Theorem 3.1,*

$$\psi = \frac{1}{\int_0^\infty S_Y(t)dt + \nu} \left\{ \int_0^\infty \exp \left(- \int_0^t r_0(x) - \int_0^t \lambda(x)dx + E_{\mathcal{H}}[a(t) + b(t)] \right) [r_0(t)E_{\mathcal{H}}[a(t)] + E_{\mathcal{H}}[\mathcal{H}a(t)b(t)]] dt \right\} \quad (4.1)$$

where $a(t) := \int_0^t e^{-\mathcal{H}v} g_W(v) m(t-v) dv$ and $b(t) := \int_0^t e^{-\mathcal{H}(t-r)} \bar{G}_W(t-r) \lambda(r) dr$.

Proof. $E[Y] = \int_0^\infty S_Y(t)dt$, where $S_Y(t)$ is given in (3.1). Hence it suffices to derive $E[M]$. The total number of jobs completed in a renewal cycle is given by $M = \sum_{j=1}^{N(Y)} \mathbb{1}_{T_j + W_j \leq Y}$. $N(Y)$ is finite for any selected renewal cycle. Hence, letting $\{(R_j, V_j)\}_{j=1}^{N(Y)}$ be a random permutation of $\{(T_j, W_j)\}_{j=1}^{N(Y)}$, $M = \sum_{j=1}^{N(Y)} \mathbb{1}_{R_j + V_j \leq Y}$ due to the mutual independence of arrival times and service times. For convenience and clarity, we introduce the following notation: $\mathfrak{R} = \{R_1, \dots, R_n\}$, $\mathfrak{V} = \{V_1, \dots, V_n\}$, $\mathfrak{H} = \{\mathcal{H}_1, \dots, \mathcal{H}_n\}$ with observed values $\mathfrak{r} = \{r_1, \dots, r_n\}$, $\mathfrak{v} = \{v_1, \dots, v_n\}$,

$\mathfrak{h} = \{\eta_{i_1}, \dots, \eta_{i_n}\}$. By Bayes' Theorem,

$$f_{\mathfrak{R}, \mathfrak{V}, \mathfrak{H}, Y, N}(\mathfrak{r}, \mathfrak{v}, \mathfrak{h}, t, n) = f_{Y|\mathfrak{R}, \mathfrak{V}, \mathfrak{H}, N}(t|\mathfrak{r}, \mathfrak{v}, \mathfrak{h}, n) f_{\mathfrak{R}, \mathfrak{V}, \mathfrak{H}, N}(\mathfrak{r}, \mathfrak{v}, \mathfrak{h}, n)$$

By A.3, the conditional distribution $f_{Y|\mathfrak{R},\mathfrak{V},\mathfrak{H},N}(t|\mathfrak{r},\mathfrak{v},\mathfrak{h},n)$ is given by

$$f_{Y|\mathfrak{R},\mathfrak{V},\mathfrak{H},N}(t|\mathfrak{r},\mathfrak{v},\mathfrak{h},n) = \exp\left(-\int_0^t r_0(s)ds - \sum_{j=1}^n \eta_j^{i_j} \min(v_j, t - r_j)\right) \left(r_0(t) + \sum_{j=1}^n \eta_j^{i_j} \mathbb{1}_{v_j > t - r_j}\right) \quad (4.2)$$

Since $\mathcal{H}_j, j = 1, \dots, n$ are i.i.d. and mutually independent of $\mathfrak{R}, \mathfrak{V}$, and N ,

$$f_{\mathfrak{R},\mathfrak{V},\mathfrak{H},N}(\mathfrak{r},\mathfrak{v},\mathfrak{h},n) = f_{\mathfrak{R},\mathfrak{V},N}(\mathfrak{r},\mathfrak{v},n) f_{\mathfrak{H}}(\mathfrak{h}) = f_{\mathfrak{R},\mathfrak{V},N}(\mathfrak{r},\mathfrak{v},n) \prod_{j=1}^n P(\mathcal{H}_j = \eta_{i_j})$$

By A.4,

$$f_{\mathfrak{R},\mathfrak{V},N}(\mathfrak{r},\mathfrak{v},n) = \frac{1}{n!} \prod_{j=1}^n \exp\left(\int_0^t \lambda(x)dx\right) \lambda(r_j) g_W(v_j)$$

and thus

$$f_{\mathfrak{R},\mathfrak{V},\mathfrak{H},N}(\mathfrak{r},\mathfrak{v},\mathfrak{h},n) = \frac{1}{n!} \prod_{j=1}^n \exp\left(-\int_0^t \lambda(s)ds\right) \lambda(r_j) g_W(v_j) P(\mathcal{H}_j = \eta_{i_j}) \quad (4.3)$$

Finally, by multiplying (4.2) and (4.3),

$$\begin{aligned} & f_{\mathfrak{R},\mathfrak{V},\mathfrak{H},Y,N(t)}(\mathfrak{r},\mathfrak{v},\mathfrak{h},t,n) \\ &= \frac{1}{n!} \left(r_0(t) + \sum_{j=1}^n \eta_{i_j} \mathbb{1}_{v_j > t - r_j}\right) \exp\left(-\int_0^t r_0(x)dx - \sum_{j=1}^n \eta_{i_j} \min(v_j, t - r_j)\right) \\ &\quad \times \prod_{j=1}^n \exp\left(-\int_0^t \lambda(x)dx\right) \lambda(r_j) g_W(v_j) P(\mathcal{H}_j = \eta_{i_j}) \\ &= \frac{1}{n!} r_0(t) \exp\left(-\int_0^t r_0(x)dx - \int_0^t \lambda(x)dx\right) \\ &\quad \times \prod_{j=1}^n \exp(-\eta_{i_j} \min(v_j, t - r_j)) \lambda(r_j) g_W(v_j) P(\mathcal{H}_j = \eta_{i_j}) \\ &\quad + \frac{1}{n!} \exp\left(-\int_0^t r_0(x)dx - \int_0^t \lambda(x)dx\right) \left(\sum_{j=1}^n \eta_{i_j} \mathbb{1}_{v_j > t - r_j}\right) \\ &\quad \times \prod_{j=1}^n \exp(-\eta_{i_j} \min(v_j, t - r_j)) \lambda(r_j) g_W(v_j) P(\mathcal{H}_j = \eta_{i_j}) \end{aligned}$$

Then

$$\begin{aligned}
E[M] &= E\left[\sum_{j=1}^{N(Y)} \mathbb{1}(R_j + V_j \leq Y)\right] \\
&= \sum_{n=1}^{\infty} \int_0^t \left[\sum_{j=1}^n \sum_{i_j=1}^m \int_0^t \int_0^{t-r_j} \int_0^t \int_0^{\infty} f_{\mathfrak{R}, \mathfrak{V}, \mathfrak{S}, Y, N(t)}(\mathfrak{r}, \mathfrak{v}, \mathfrak{h}, t, n) d\mathfrak{v}_{-j} d\mathfrak{r}_{-j} dv_j dr_j \right] dt \\
&= \sum_{n=1}^{\infty} \int_0^{\infty} \frac{1}{n!} r_0(t) \exp\left(-\int_0^t r_0(x) dx - \int_0^t \lambda(x) dx\right) \\
&\quad \times n E_{\mathcal{H}} \left[\int_0^t \int_0^{t-r} \exp(-\mathcal{H}v) g_W(v) dv \lambda(r) dr \right] \\
&\quad \times \left\{ E_{\mathcal{H}} \left[\int_0^t \int_0^{t-r} \exp(-\mathcal{H}v) g_W(v) dv \lambda(r) dr + \int_0^t \exp(-\mathcal{H}(t-r)) \bar{G}_W(t-r) \lambda(r) dr \right] \right\}^{n-1} dt \\
&+ \sum_{n=1}^{\infty} \int_0^{\infty} \frac{1}{n!} \exp\left(-\int_0^t r_0(x) dx - \int_0^t \lambda(x) dx\right) \\
&\quad \times n(n-1) E_{\mathcal{H}} \left[\mathcal{H} \int_0^t \int_0^{t-r} \exp(-\mathcal{H}v) g_W(v) dv \lambda(r) dr \int_0^t \exp(-\mathcal{H}(t-r)) \bar{G}_W(t-r) \lambda(r) dr \right] \\
&\quad \times \left\{ E_{\mathcal{H}} \left[\int_0^t \int_0^{t-r} \exp(-\mathcal{H}v) g_W(v) dv \lambda(r) dr + \int_0^t \exp(-\mathcal{H}(t-r)) \bar{G}_W(t-r) \lambda(r) dr \right] \right\}^{n-2} dt
\end{aligned}$$

Let $a(t) = \int_0^t \int_0^{t-r} \exp(-\mathcal{H}v) g_W(v) dv \lambda(r) dr$. Through a change of variables,

$$a(t) = \int_0^t \exp(-\mathcal{H}v) g_W(v) m(t-v) dv$$

Denoting $b(t) = \int_0^t \exp(-\mathcal{H}(t-r)) \bar{G}_W(t-r) \lambda(r) dr$, the above simplifies to

$$\begin{aligned}
E[M] &= \sum_{n=1}^{\infty} \int_0^{\infty} \frac{1}{(n-1)!} r_0(t) E_{\mathcal{H}}[a(t)] (E_{\mathcal{H}}[a(t) + b(t)])^{n-1} \exp\left(-\int_0^t r_0(x) dx - \int_0^t \lambda(x) dx\right) dt \\
&\quad + \sum_{n=2}^{\infty} \int_0^{\infty} \frac{1}{(n-2)!} E_{\mathcal{H}}[\mathcal{H}a(t)b(t)] (E_{\mathcal{H}}[a(t) + b(t)])^{n-2} dt \\
&= \int_0^{\infty} r_0(t) \exp\left(-\int_0^t r_0(x) dx - \int_0^t \lambda(x) dx\right) E_{\mathcal{H}}[a(t)] \left(\sum_{n=1}^{\infty} \frac{1}{(n-1)!} (E_{\mathcal{H}}[a(t) + b(t)])^{n-1} \right) dt \\
&\quad + \int_0^{\infty} \exp\left(-\int_0^t r_0(x) dx - \int_0^t \lambda(x) dx\right) E_{\mathcal{H}}[\mathcal{H}a(t)b(t)] \left(\sum_{n=2}^{\infty} \frac{1}{(n-2)!} (E_{\mathcal{H}}[a(t) + b(t)])^{n-2} \right) dt \\
&= \int_0^{\infty} r_0(t) \exp\left(-\int_0^t r_0(x) dx - \int_0^t \lambda(x) dx\right) E_{\mathcal{H}}[a(t)] \exp(E_{\mathcal{H}}[a(t) + b(t)]) dt \\
&\quad + \int_0^{\infty} \exp\left(-\int_0^t r_0(x) dx - \int_0^t \lambda(x) dx\right) E_{\mathcal{H}}[\mathcal{H}a(t)b(t)] \exp(E_{\mathcal{H}}[a(t) + b(t)]) dt \\
&= \int_0^{\infty} \exp\left(-\int_0^t r_0(x) dx - \int_0^t \lambda(x) dx + E_{\mathcal{H}}[a(t) + b(t)]\right) [r_0(t) E_{\mathcal{H}}[a(t)] + E_{\mathcal{H}}[\mathcal{H}a(t)b(t)]] dt
\end{aligned}$$

Finally, we see that

$$\psi = \frac{1}{\int_0^{\infty} S_Y(t) dt + \nu} \left\{ \int_0^{\infty} \exp\left(-\int_0^t r_0(x) dx - \int_0^t \lambda(x) dx + E_{\mathcal{H}}[a(t) + b(t)]\right) [r_0(t) E_{\mathcal{H}}[a(t)] + E_{\mathcal{H}}[\mathcal{H}a(t)b(t)]] dt \right\}$$

5. Conclusion

This generalization provides a more realistic model to any server under nonconstant traffic. By specifying the service life distribution, the intensity function, and the distribution of the stresses, one can model almost any traffic situation. Future work will include investigating correlated Poisson streams, a partitioned server, and sufficient conditions for the maximum of the efficiency ψ .

Appendix A.

Lemma A.1. *Let $\{N(t)\}$ be a nonhomogeneous Poisson process describing the arrival of client requests to the web server with intensity $\lambda(t)$, and let $\{T_j\}_{j=1}^{N(t)}$ be the arrival times of the client requests. Then, given $N(t) = n$, the conditional joint distribution of is equal to the joint distribution of the order statistics $T'_{[1]}, \dots, T'_{[n]}$, where T'_1, \dots, T'_n are i.i.d. with pdf $f(x) = \frac{\lambda(x)}{m(t)}$, where $m(t) = \int_0^t \lambda(s)ds$ Formally,*

$$f_{T_1, \dots, T_n | N(t)=n}(t_1, \dots, t_n | n) = f_{T'_{[1]}, \dots, T'_{[n]}}(t_1, \dots, t_n) = \frac{n!}{m(t)} \prod_{i=1}^n \lambda(t_i)$$

for $0 \leq t_1 \leq \dots \leq t_n \leq t$.

Proof. Let $N(t) = n$, and let $0 \leq t_1 \leq \dots \leq t_n \leq t$. Let $h_i, i = 1, \dots, n$ be small enough such that $t_i + h_i < t_{i+1} \forall i = 1, \dots, n-1$. Let A_i be the event where the server sees exactly one arrival in $[t_i, t_i + h_i]$, and B be the event that no events arrive outside the set $U = [0, t_1] \cup [t_1, t_1 + h_1] \cup [t_2, t_2 + h_2] \cup \dots \cup [t_n, t_n + h_n] \cup [t_n + h_n, t]$. Then

$$P(t_i \leq T_i \leq t_i + h_i, i = 1, \dots, n | N(t) = n) = \frac{P(A_1 \cap A_2 \cap \dots \cap A_n \cap B)}{P(N(t) = n)}$$

By definition, $P(N(t) = n) = \frac{e^{-m(t)} m(t)^n}{n!}$. For each $i = 1, \dots, n$,

$$\begin{aligned} P(t_i \leq T_i \leq t_i + h_i) &= P(N(t_i + h_i) - N(t_i) = 1) \\ &= e^{-(m(t_i + h_i) - m(t_i))} (m(t_i + h_i) - m(t_i)) \end{aligned}$$

The complement of the set U can be broken into the disjoint intervals $[0, t_1], [t_1 + h_1, t_2], \dots, [t_i + h_i, t_{i+1}], \dots, [t_n + h_n, t]$. Then

$$\begin{aligned} P(B) &= P(N(t) - N(t_n + h_n) = 0) P(N(t_1) - N(0) = 0) \prod_{i=1}^{n-1} P(N(t_{i+1}) - N(t_i + h_i) = 0) \\ &= e^{-m(t_1)} e^{-(m(t) - m(t_n + h_n))} \prod_{i=1}^{n-1} e^{-(m(t_{i+1}) - m(t_i + h_i))} \\ &= \exp \left(- \left[m(t) + \sum_{i=1}^n m(t_i) - \sum_{i=1}^n m(t_i + h_i) \right] \right) \end{aligned}$$

Now,

$$\begin{aligned}
P(t_i \leq T_i \leq t_i + h_i, i = 1, \dots, n | N(t) = n) &= \prod_{i=1}^n P(t_i \leq T_i \leq t_i + h_i) \\
&= \frac{\prod_{i=1}^n [m(t_i + h_i) - m(t_i)] e^{-\sum_{i=1}^n [m(t_i + h_i) - m(t_i)]} e^{-[m(t) + \sum_{i=1}^n m(t_i) - \sum_{i=1}^n m(t_i + h_i)]}}{\frac{e^{-m(t)} m(t)^n}{n!}} \\
&= n! \prod_{i=1}^n \frac{m(t_i + h_i) - m(t_i)}{m(t)} \\
&= n! \prod_{i=1}^n \frac{\lambda(t_i)}{m(t)}
\end{aligned}$$

by letting $h_i \rightarrow 0 \forall i$.

Lemma A.2.

$$E \left[E \left[\exp \left(-\eta_j^{i_{j'}} \min(W_j, t - T_j') | W_j \right) \right] \right] = \frac{1}{m(t)} \left(m(t) - \eta_j^{i_{j'}} \int_0^t \exp(-\eta_j^{i_{j'}} w) m(t - w) \bar{G}_W(w) dw \right) \quad (\text{A.1})$$

Proof. Two cases are considered: (1): $w \leq t$, and (2): $w > t$. For $w \leq t$,

$$\begin{aligned}
E \left[\exp(-\eta_{i_j} \min(W_j, t - T_j')) | W_j = w \right] &= \int_0^{t-w} \exp(-\eta_{i_j} x) \frac{\lambda(x)}{m(t)} dx + \int_{t-w}^t \exp(-\eta_{i_j} (t - x)) \frac{\lambda(x)}{m(t)} dx \\
&= \exp(-\eta_{i_j} w) \frac{m(t - w)}{m(t)} + \exp(-\eta_{i_j} t) \int_0^t \exp(\eta_{i_j} x) \frac{\lambda(x)}{m(t)} dx
\end{aligned}$$

For $w > t$,

$$E \left[\exp(-\eta_{i_j} \min(W_j, t - T_j')) | W_j = w \right] = \exp(-\eta_{i_j} t) \int_0^t \exp(\eta_{i_j} x) \frac{\lambda(x)}{m(t)} dx$$

Therefore,

$$\begin{aligned}
E_W \left[\exp(-\eta_{i_j} \min(W_j, t - T_j')) \right] &= E_W \left[E \left[\exp(-\eta_{i_j} \min(W_j, t - T_j')) | W_j = w \right] \right] \\
&= \frac{1}{m(t)} \left(\int_0^t \exp(-\eta_{i_j} w) m(t - w) g_W(w) dw \right. \\
&\quad \left. + \exp(-\eta_{i_j} t) \int_0^t \int_{t-w}^t \exp(\eta_{i_j} x) \lambda(x) dx g_W(w) dw \right. \\
&\quad \left. + \bar{G}_W(t) \exp(-\eta_{i_j} t) \int_0^t \exp(\eta_{i_j} x) \lambda(x) dx \right)
\end{aligned}$$

Focusing on the second term, we make the change of variables $w = t - x$ and change the order of intergration, yielding a new second term.

$$\begin{aligned}
&= \frac{1}{m(t)} \left(\int_0^t \exp(-\eta_{i_j} w) m(t-w) g_W(w) dw \right. \\
&\quad \left. + \exp(-\eta_{i_j} t) \int_0^t \exp(\eta_{i_j} x) \lambda(x) \int_{t-x}^t g_W(w) dw dx \right. \\
&\quad \left. + \bar{G}_W(t) \exp(-\eta_{i_j} t) \int_0^t \exp(\eta_{i_j} x) \lambda(x) dx \right)
\end{aligned}$$

Combining the second and third terms:

$$\begin{aligned}
&= \frac{1}{m(t)} \left(\int_0^t \exp(-\eta_{i_j} w) m(t-w) g_W(w) dw \right. \\
&\quad \left. + \exp(-\eta_{i_j} t) \int_0^t \exp(\eta_{i_j} x) \lambda(x) \bar{G}_w(t-x) dx \right)
\end{aligned}$$

Changing variables again in the second term, using $w = t - x$, we get

$$\begin{aligned}
&= \frac{1}{m(t)} \left(\int_0^t \exp(-\eta_{i_j} w) m(t-w) g_W(w) dw \right. \\
&\quad \left. + \int_0^t \exp(-\eta_{i_j} w) \lambda(t-w) \bar{G}_w(w) dw \right)
\end{aligned}$$

Integrating the first term by parts, we get

$$\begin{aligned}
&= \frac{1}{m(t)} \left(\left[-\exp(-\eta_{i_j} w) m(t-w) \bar{G}_w(w) \right]_0^t \right. \\
&\quad \left. - \int_0^t (\eta_{i_j} \exp(-\eta_{i_j} w) m(t-w) + \exp(-\eta w) \lambda(t-w)) \bar{G}_W(w) dw \right. \\
&\quad \left. + \int_0^t \exp(-\eta_{i_j} w) \lambda(t-w) \bar{G}_w(w) dw \right) \\
&= \frac{1}{m(t)} \left(m(t) - \eta_{i_j} \int_0^t \exp(-\eta_{i_j} w) m(t-w) \bar{G}_W(w) dw \right)
\end{aligned}$$

Lemma A.3. *Under all conditions of Theorem 4.1, the conditional distribution $f_{Y|\mathfrak{R}, \mathfrak{V}, \mathfrak{H}, N}(t|\mathfrak{r}, \mathfrak{v}, \mathfrak{h}, n)$ is given by*

$$f_{Y|\mathfrak{R}, \mathfrak{V}, \mathfrak{H}, N}(t|\mathfrak{r}, \mathfrak{v}, \mathfrak{h}, n) = \exp \left(- \int_0^t r_0(s) ds - \sum_{j=1}^n \eta_{i_j} \min(v_j, t - r_j) \right) \left(r_0(t) + \sum_{j=1}^n \eta_{i_j} \mathbb{1}_{v_j > t - r_j} \right)$$

Proof. Let $C = \{\mathfrak{R} = \mathfrak{r}, \mathfrak{V} = \mathfrak{v}, \mathfrak{H} = \mathfrak{h}, N(t) = n\}$. Then

$$f_{Y|C}(t|c) = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} (P(Y > t|C = c) - P(Y > t + \Delta t|C = c)) \quad (\text{A.2})$$

Recall $\mathcal{B}(s) = r_0(s) + \sum_{j=1}^{N(t)} \mathbb{I}(R_j \leq s \leq R_j + V_j)$. It remains to derive $P(Y > t + \Delta t | C = c)$. Let $A_k = \{N(t + \Delta t) - N(t) = k\}$ be the event such that k requests arrived between t and $t + \Delta t$. From the definition of a nonhomogenous Poisson process ([2]),

$$\begin{aligned} P(A_0) &= 1 - \lambda(t)\Delta t + o(\Delta t) = e^{-(m(t+\Delta t)-m(t))} \\ P(A_1) &= \lambda(t)\Delta t + o(\Delta t) = (m(t+\Delta t) - m(t))e^{-(m(t+\Delta t)-m(t))} \\ &\vdots \\ P(A_k) &= \frac{(m(t+\Delta t) - m(t))^k}{k!} e^{-(m(t+\Delta t)-m(t))} \end{aligned}$$

for $k \geq 2$. By the Law of Total Probability,

$$P(Y > t + \Delta t | C = c) = \sum_{k=0}^{\infty} P(Y > t + \Delta t | C \cap A_k) P(A_k) \quad (\text{A.3})$$

(A.3) will be computed for each fixed k , and it will be shown that for $k \geq 2$, (A.3) is of order $o(\Delta t)$. For $k = 0$,

$$P(Y > t + \Delta t | C \cap A_0) P(A_0) = e^{-\int_0^{t+\Delta t} \mathcal{B}(s) ds} (1 - \lambda(t)\Delta t + o(\Delta t)) = e^{-\int_0^t \mathcal{B}(s) ds} e^{-\int_t^{t+\Delta t} \mathcal{B}(s) ds} (1 - \lambda(t)\Delta t + o(\Delta t)) \quad (\text{A.4})$$

where

$$\begin{aligned} \int_t^{t+\Delta t} \mathcal{B}(s) ds &= \int_t^{t+\Delta t} r_0(s) ds + \left(\sum_{j=1}^n \eta_{i_j} \mathbb{1}_{v_j > t-r_j} \right) \Delta t \\ &= r_0(t) + o(\Delta t) + \left(\sum_{j=1}^n \eta_{i_j} \mathbb{1}_{v_j > t-r_j} \right) \Delta t \end{aligned}$$

Using the fact that $e^{-x+o(x)} = 1 - x + o(x)$ for small x ,

$$e^{-\int_t^{t+\Delta t} \mathcal{B}(s) ds} = 1 - \left(r_0(t) + \sum_{j=1}^n \eta_{i_j} \mathbb{1}_{v_j > t-r_j} \right) \Delta t + o(\Delta t)$$

Substituting into (A.4),

$$\begin{aligned} P(Y > t + \Delta t | C \cap A_0) P(A_0) &= \left[e^{-\int_0^t \mathcal{B}(s) ds} 1 - \left(r_0(t) + \sum_{j=1}^n \eta_{i_j} \mathbb{1}_{v_j > t-r_j} \right) \Delta t + o(\Delta t) \right] [1 - \lambda(t)\Delta t + o(\Delta t)] \\ &= e^{-\int_0^t \mathcal{B}(s) ds} \left(1 - [r_0(t) + \sum_{j=1}^n \eta_{i_j} \mathbb{1}_{v_j > t-r_j}] \Delta t - \lambda(t)\Delta t + o(\Delta t) \right) \quad (\text{A.5}) \end{aligned}$$

for $k = 1$, one request has arrived in $[t, t + \Delta t]$ in addition to the n fixed arrivals given by C . Call this arrival t_1 , with corresponding time to completion w_1 . We then have two cases:

(1) $t_1 + w_1 < t + \Delta t$, that is, the request arrives and is serviced before $t + \Delta t$, or

(2) $t_1 + w_1 > t + \Delta t$. In this case, the service time is greater than Δt .

With both of these cases, the failure rate increases by an $\eta_{i_{t_1}}$ for a time smaller than Δt , which we will denote as $\Delta^1(t)$. Then

$$\Delta^1(t) = \begin{cases} w_1, & t_1 + w_1 < t + \Delta t \\ t + \Delta t - t_1, & t_1 + w_1 > t + \Delta t \end{cases}$$

Applying similar logic from earlier to calculate $P(Y > t + \Delta t | C \cap A_1)P(A_1)$,

$$\exp\left(-\int_t^{t+\Delta t} \mathcal{B}(s)ds\right) = \exp\left(-r_0(t) + \left[\sum_{j=1}^n \eta_{i_j} \mathbb{1}_{v_j > t-r_j}\right]\Delta t + \eta_{t_1}^{i_{t_1}} \Delta^1 t + o(\Delta t)\right)$$

which simplifies to

$$\exp\left(-\int_t^{t+\Delta t} \mathcal{B}(s)ds\right) = 1 - [r_0(t) + \sum_{j=1}^n \eta_{i_j} \mathbb{1}_{v_j > t-r_j}]\Delta t + \eta_{t_1}^{i_{t_1}} \Delta^1 t + o(\Delta t)$$

Now,

$$\begin{aligned} P(Y > t + \Delta t | C \cap A_1)P(A_1) &= e^{-\int_0^t \mathcal{B}(s)ds} \left([1 - [r_0(t) + \sum_{j=1}^n \eta_{i_j} \mathbb{1}_{v_j > t-r_j}]\Delta t + \eta_{t_1}^{i_{t_1}} \Delta^1 t + o(\Delta t)] [\lambda(t)\Delta t + o(\Delta t)] \right) \\ &= e^{-\int_0^t (\lambda(s)\Delta t + o(\Delta t))} \end{aligned} \quad (\text{A.6})$$

For $k \geq 2$, it will be shown that the contribution to (A.3) are negligible.

$$P(Y > t + \Delta t | C \cap A_k) = e^{-\int_0^t \mathcal{B}(s)ds} \left[1 - [(r_0(t) + \sum_{j=1}^n \eta_{i_j} \mathbb{1}_{v_j > t-r_j})\Delta t + \sum_{l=1}^k \eta_{t_l}^{i_{t_l}} \Delta^l t] + o(\Delta t) \right]$$

Hence $\forall k \geq 2, P(Y > t + \Delta t | C \cap A_k) \leq e^{-\int_0^t \mathcal{B}(s)ds}$ and thus

$$\begin{aligned} \sum_{k=2}^{\infty} P(Y > t + \Delta t | C \cap A_k)P(A_k) &\leq e^{-\int_0^t \mathcal{B}(s)ds} \sum_{k=2}^{\infty} P(A_k) \\ &= e^{-\int_0^t \mathcal{B}(s)ds} (1 - P(A_0) - P(A_1)) \\ &= e^{-\int_0^t \mathcal{B}(s)ds} o(\Delta t) \end{aligned}$$

Therefore,

$$P(Y > t + \Delta t | C) = e^{-\int_0^t \mathcal{B}(s)ds} \left(1 - (r_0(t) + \sum_{j=1}^n \eta_{i_j} \mathbb{1}_{v_j > t-r_j})\Delta t \right) + o(\Delta t)$$

and

$$\frac{1}{\Delta t} (P(Y > t | C = c) - P(Y > t + \Delta t | C = c)) = e^{-\int_0^1 \mathcal{B}(s)ds} \left(r_0(t) + \sum_{j=1}^n \eta_{i_j} \mathbb{1}_{v_j > t-r_j} + \frac{o(\Delta t)}{\Delta t} \right)$$

Then by letting $\Delta t \rightarrow 0$,

$$f_{Y|\mathfrak{R}, \mathfrak{V}, \mathfrak{S}, N}(t|\mathfrak{r}, \mathfrak{v}, \mathfrak{h}, n) = \exp \left(- \int_0^t r_0(s) ds - \sum_{j=1}^n \eta_{i_j} \min(v_j, t - r_j) \right) \left(r_0(t) + \sum_{j=1}^n \eta_{i_j} \mathbf{1}_{v_j > t - r_j} \right)$$

Lemma A.4.

$$f_{\mathfrak{R}, \mathfrak{V}, N(t)}(\mathfrak{r}, \mathfrak{v}, n) = \frac{1}{n!} \exp \left(- \int_0^t \lambda(s) ds \right) \prod_{j=1}^n \lambda(r_j) g_W(v_j)$$

Proof. Let $\mathcal{W} = (W_1, \dots, W_n)$ be the service times under the condition that $N(t) = n$, and let $\mathcal{W} = \underline{w} = (w_1, \dots, w_n)$. Let $(r_{[1]}, \dots, r_{[n]})$ be the ordered vector of \mathfrak{r} . There are $n!$ possible orderings of \mathfrak{r} . Now, $P(r_{[i]} \leq R_{[i]} \leq r_{[i]} + h_i)$ for some $h_i > 0, i = 1, \dots, n$ is given by

$$P(r_{[i]} \leq R_{[i]} \leq r_{[i]} + h_i) = e^{-(m(r_{[i]} + h_i) - m(r_{[i]}))} (m(r_{[i]} + h_i) - m(r_{[i]}))$$

In a similar manner to the proof of Lemma A.1, we may see that the joint distribution of \mathfrak{R} is identical to the distribution of the order statistics of \mathfrak{R} :

$$f_{\mathfrak{R}, N(t)}(\mathfrak{r}, n) = \frac{1}{n!} \prod_{j=1}^n \lambda(r_j) \exp \left(- \int_0^t \lambda(s) ds \right)$$

$\mathfrak{R}, \mathfrak{V}$ are mutually independent. Therefore,

$$f_{\mathfrak{R}, \mathfrak{V}, N(t)}(\mathfrak{r}, \mathfrak{v}, n) = \frac{1}{n!} \exp \left(- \int_0^t \lambda(s) ds \right) \prod_{j=1}^n \lambda(r_j) g_W(v_j)$$

Acknowledgements

My sincerest thanks to Andrzej Korzeniowski and Jason Hathcock for technical assistance and advice.

References

- [1] , CHA, KI HWAN AND LEE, EUI YONG (2011). A Stochastic Breakdown Model for an Unreliable Web Server System and an Optimal Admission Control Policy., *Journal of Applied Probability*. **48**(2), 453-466.
- [2] ROSS, S. (1996). *Stochastic Processes*, 2nd edn. John Wiley, UK.